

# PC5213 Biophysics Final Report

## Gillespie Monte Carlo Algorithm for Steady State Kinetics of Enzyme Reaction

Xin Chen A0151856Y

May 7, 2017

Based on a simple system, a retrospective summary on steady-state kinetics of enzyme-catalysed chemical reactions is firstly composed as the introduction. Next, the second section is a review of analytic methods and Gillespie Monte Carlo algorithm for studying the steady-state kinetics of the system, which can lead to the Michaelis Menten theory. The equivalence of the direct and "first-reaction" Gillespie Monte Carlo methods is shown in the second section as well. Finally, in the last section, simulation results from both of the two Gillespie methods are demonstrated and compared.

## 1 Introduction

Enzymes can catalyze chemical reactions by lowering the energy barrier which is needed to be crossed over during the reaction. A general enzyme-catalysed reaction is described as the following form,



which involves an enzyme E, a substrate S and a product P. In this reaction, E and S firstly bind together to form a complex ES, and ES then releases the product P. Here  $k_1$ ,  $k_2$  and  $k_3$  are the corresponding first-order rate constants. If denoting the concentrations of each component as  $C_E$ ,  $C_S$ ,  $C_{ES}$  and  $C_P$ , the dynamics of this reaction is formulated by the following master equation.

$$\frac{d}{dt} \begin{pmatrix} C_E \\ C_{ES} \\ C_S \\ C_P \end{pmatrix} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \quad (2)$$

where the matrix is called stoichiometric coefficient matrix and the entities of its three columns are corresponding to stoichiometric coefficients of the three reactions. And  $v_1$ ,  $v_2$ ,  $v_3$  denote the reaction speed, i.e.

$$\begin{aligned} v_1 &= k_1 C_S C_E \\ v_2 &= k_2 C_{ES} \\ v_3 &= k_3 C_{ES} \end{aligned} \quad (3)$$

In a typical enzyme-catalysed reaction, one may assume that  $C_S$ ,  $C_P$  are much higher than  $C_E$  and  $C_{ES}$ , which results in the fact that  $C_S$  and  $C_P$  are approximately constants during a long-run. Thus the states of enzyme E (either as empty E or as the complex ES) are considered as the major factor. This allows us to reduce the model by the following steps. First, omit the two rows of  $C_S$  and  $C_P$  in equation (2). Next, divide  $C_0 = C_E + C_{ES}$  on both sides of equation (2), and further define the pseudo-first order rate constant  $\tilde{k}_1 = k_1 C_S$ . Then the original equation (2) can be reduced to

$$\frac{d}{dt} \begin{pmatrix} p_E \\ p_{ES} \end{pmatrix} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \tilde{k}_1 p_E \\ k_2 p_{ES} \\ k_3 p_{ES} \end{pmatrix}, \quad (4)$$

or

$$\frac{d}{dt} \begin{pmatrix} p_E \\ p_{ES} \end{pmatrix} = \begin{pmatrix} -\tilde{k}_1 & k_2 + k_3 \\ \tilde{k}_1 & -k_2 - k_3 \end{pmatrix} \begin{pmatrix} p_E \\ p_{ES} \end{pmatrix}. \quad (5)$$

Here the  $p_i = C_i/C_0$ ,  $i = E, ES$  is the probability that the enzyme is at state E or ES. This reduced dynamic system can be shown in figure 1.

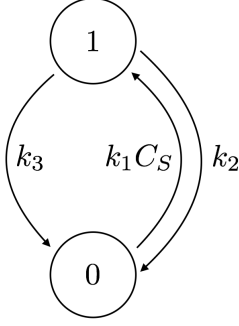


Figure 1: dynamic graph of the system. state 0 = E and 1 = ES

Next we consider the situation when the system reaches a steady state, i.e. the l.h.s of equation (5) is zero. Notice that the normalization condition

$$p_E + p_{ES} = 1 \quad (6)$$

should also be satisfied. So now we have three equations for only two variables, which means there is one redundant equation. To form a soluable linear system, we simply cancel the second row in equation (5), then add the normalization condition into it. Finally the equation for the steady state becomes

$$\begin{pmatrix} -\tilde{k}_1 & k_2 + k_3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} p_E \\ p_{ES} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (7)$$

To understand the kinetics of the steady state, we can simply solve the equation (7), or use Gillespie Monte Carlo algorithm. These two methods will be discussed in the next section.

## 2 Methods

### 2.1 Analytic Solution

One straightforward method to solve the steady state equation (7) analytically is Cramer's rule. The solution in this case simply is

$$p_E = \frac{k_2 + k_3}{\tilde{k}_1 + k_2 + k_3} \quad (8)$$

$$p_{ES} = \frac{\tilde{k}_1}{\tilde{k}_1 + k_2 + k_3}.$$

Alternatively, the solution can be written out by diagram method [3]. To apply this method, one need firstly construct partial graphs, which are undirected graphs containing the maximum number of edges without forming any cycles. In this case, there are two of them, as shown in figure 2. Then, by adding arrows into the partial graphs, directed graphs are formed. There are three of them in this case, as shown in figure 3.

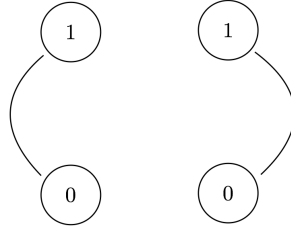


Figure 2: partial graphs of figure 1

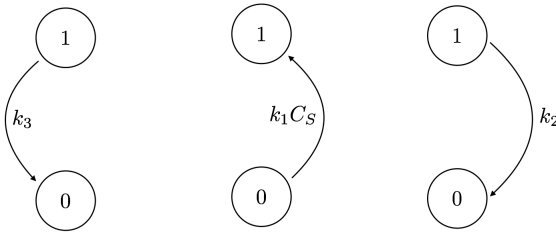


Figure 3: directed graphs following the partial graphs

Each graph in figure 3 is corresponding to a rate constant (or a product of rate constants in a general case), Then the state probabilities are

$$p_i = \frac{\text{sum of directed graphs with state } i \text{ being pointed to}}{\text{sum of all directed graphs}}. \quad (9)$$

By calculating so, equation (8) will be recovered exactly.

A major concern here would be the product rate of this reaction regime. The product rate in this case can be described by the probability flux from state 1 to 0 through the left path with the rate constant  $k_3$ ,

$$J_{1 \rightarrow 0}^{k_3} = k_3 p_{ES} = \frac{k_3 \tilde{k}_1}{\tilde{k}_1 + k_2 + k_3}. \quad (10)$$

$J_{1 \rightarrow 0}^{k_3}$  means, for one enzyme molecule E, the average number of reactions with rate constant  $k_3$  which take place per unit time, i.e. the product rate with one enzyme in the system. If we consider the total product rate  $v_P$  in a solution, the initial concentration of the enzymes should be multiplied. Then the Michaelis Menten equation [4] is recovered.

$$v_P = C_0 J_{1 \rightarrow 0}^{k_3} = \frac{v_{\max} C_S}{C_S + K_M}, \quad (11)$$

where  $v_{\max} = k_3 C_0$  denotes the maximum product rate and  $K_M = k_2 + k_3/k_1$  is the Michaelis constant. This analytic result will be compared with Gillespie Monte Carlo simulation result in the last section.

## 2.2 Gillespie Monte Carlo Algorithm

The core idea of Gillespie Monte Carlo algorithm [1, 2] is to characterize the kinetic properties of the system by *reaction probability density*  $\rho(\mu, \tau)$ , where  $\mu$  denotes all possible transitions with respect to current state, and  $\tau$  is a positive quantity called transition time. In this regime,  $\rho(\mu, \tau) d\tau$  means the probability that the very first reaction that takes place in the infinitesimal interval  $(t + \tau, t + \tau + d\tau]$  is the  $\mu$ th reaction. That is, more explicitly, the probability that no reaction takes place from current time  $t$  until  $t + \tau$ , while the  $\mu$ th reaction then takes place in the next  $d\tau$  time, as illustrated in figure 4.

Here comes a brief derivation for  $\rho(\mu, \tau)$ . According to the abovementioned,  $\rho(\mu, \tau)$  is a joint probability density of the two events shown in figure 4. The probability that  $\mu$ th reaction takes place in  $(t + \tau, t + \tau + d\tau]$  simply is, by definition of rate constant,  $k_\mu d\tau$ , and the probability of the other event can be formulated by the following deduction. Firstly divide the interval  $(t, t + \tau]$  into  $M$  segments whose length of each is a very small number  $\varepsilon = \tau/M$ . Secondly, consider the probability that  $\nu$ th reaction does not take place in interval  $\varepsilon$ , which should be  $1 - k_\nu \varepsilon$ . This implies

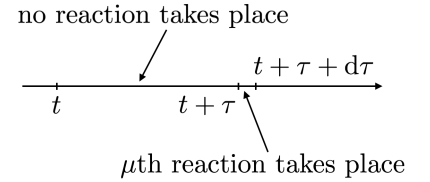


Figure 4: explanation of  $\rho(\mu, \tau)$

$$\begin{aligned} & p(\text{no reaction takes place in next } \tau) \\ &= \lim_{M \rightarrow \infty} \left[ 1 - \sum_{\nu} k_{\nu} \frac{\tau}{M} \right]^M = e^{-\sum_{\nu} k_{\nu} \tau} = e^{-K\tau}, \end{aligned} \quad (12)$$

where the summation is over all possible reactions at current time and  $K$  denotes the summation for simplicity. So finally, the reaction probability density is

$$\rho(\mu, \tau) = k_{\mu} e^{-K\tau}. \quad (13)$$

Based on the above knowledge, Gillespie Monte Carlo can be understood without difficulties. The basic idea is to simulate the transition process by continuously generating random variable pairs  $(\mu, \tau)$  according the reaction probability density  $\rho(\mu, \tau)$ . In order to do so, one need calculate the two marginal probabilities of  $\rho(\mu, \tau)$  as follows.

$$\begin{aligned} \rho_0(\tau) &= \sum_{\nu} k_{\nu} \rho(\mu, \tau) = K e^{-K\tau} \\ p_0(\mu) &= \int_0^{+\infty} \rho(\mu, \tau) d\tau = \frac{k_{\mu}}{K}, \end{aligned} \quad (14)$$

Thus the transition time  $\tau$  is exponentially distributed, and reaction indexes  $\mu$  follows a discrete distribution which is proportional to  $k_\mu$ . As can be seen, the multiplication of the two marginal probabilities recovers the reaction distribution density, which implies the following two events A and B are independent with each other.

A: the next transition is the  $\mu$ th transition.

B: the next transition will take place in time  $\tau$ .

Therefore the pair  $(\mu, \tau)$  can be sampled separately according to the following formulas.

$$\tau = \frac{1}{K} \ln \frac{1}{r_1}, \quad \sum_{\nu=1}^{\mu-1} k_\nu < r_2 K \leq \sum_{\nu=1}^{\mu} k_\nu \quad (15)$$

$r_1, r_2 \in (0, 1]$  : uniformly distributed random variables.

Then the direct procedure of Gillespie Monte Carlo can be described as follows.

1. Initialization: set  $t = 0$  and state  $= i$  (0 or 1 in the system described by figure 1).
2. Sampling: generator random pair  $(\mu, \tau)$  by using equation (15)
3. Update: update current time  $t$  by adding  $\tau$  and current state by triggering the  $\mu$ th reaction.
4. Repeat step 2 and 3 for enough times.
5. Compute the life-span time fractions for states of interest, and regard the fractions as state probabilities.

Alternatively, one can use another procedure called "first-reaction" method. In this method, one determines the next reaction and transition time at one time by sampling and comparing the quantities

$$\tau_\mu = \frac{1}{k_\mu} \ln \left( \frac{1}{r} \right), \quad (16)$$

where  $\mu$  varies among all possible transitions with respect to current state, and  $r \in (0, 1]$  is a uniformly distributed random number. Once these quantities are computed, the shortest  $\tau_{\text{shortest}}$  wins and the corresponding transition will be the "first reaction". Then the time will be updated by adding  $\tau_{\text{shortest}}$  and the state will be updated according the winning transition.

Both of these two procedures will be implemented and their results will be compared in the last section, while here I try to show the equivalence of them. The corresponding probability density of the sampling method shown in equation (16) is

$$\rho_\mu(\tau) = k_\mu e^{-k_\mu \tau}. \quad (17)$$

Notice the difference between  $\rho_\mu(\tau)$  here and  $\rho_0(\tau)$  in equation (14) is that  $\rho_0$  refers to the global transition time while  $\rho_\mu$  refers to the transition  $\mu$  only. For simplicity, the system shown in figure 1 is taken as an example for following derivation without losing any generality. Suppose the system is currently at state 1, then the two possible transitions are  $k_2$  and  $k_3$ . The corresponding transition time probability densities of them are denoted as

$$\begin{aligned} \rho_2(\tau) &= k_2 e^{-k_2 \tau} \\ \rho_3(\tau) &= k_3 e^{-k_3 \tau}. \end{aligned} \quad (18)$$

The meaning of them, taking  $\rho_2$  as an example, is that  $\rho_2(\tau) d\tau$  is the probability that the transition  $k_2$  will take place in the next  $d\tau$ , terminating a period with no  $k_2$  taking place.

The following proof is divided into two parts: the equivalence of the global transition time distribution, and the equivalence of the discrete reaction index distribution. To show the first equivalence, one may ask what is the probability that any one transition of the two will take place in the next  $d\tau$ . Explicitly, this is the probability of the following event:

$k_2$  will take place in the  $(t + \tau, t + \tau + d\tau]$  AND  $k_3$  does not in  $(t, t + \tau]$ ,  
OR  $k_3$  will take place in the  $(t + \tau, t + \tau + d\tau]$  AND  $k_2$  does not in  $(t, t + \tau]$ .

Then this probability can be written down directly (recall equation (12)).

$$\rho_2(\tau) d\tau \times e^{-k_3\tau} + \rho_3(\tau) d\tau \times e^{-k_2\tau} = (k_2 + k_3)e^{-(k_2+k_3)\tau} d\tau \quad (19)$$

Therefore, this recovers the form of  $\rho_0$  in equation (14). Thus, the global transition time distribution in direct method and "first-reaction" method are the same. The second equivalence can be shown by asking, in the regime of "first-reaction" method, what is the probability that the reaction which will firstly take place subsequently is the  $\mu$ th reaction. Taking  $\mu = 2$  as an example, this requires us to calculate the probability that  $\tau_2 < \tau_3$  ( $\tau_2$  and  $\tau_3$  refer to the quantities in equation (16)). This is also a straightforward calculation as follows.

$$\begin{aligned} p(\tau_2 < \tau_3) &= \int_0^{+\infty} d\tau_3 \int_0^{\tau_3} d\tau_2 \rho(\tau_2, \tau_3) \\ &= \int_0^{+\infty} d\tau_3 \int_0^{\tau_3} d\tau_2 k_2 e^{-k_2\tau_2} k_3 e^{-k_3\tau_3} \\ &= \frac{k_2}{k_2 + k_3}. \end{aligned} \quad (20)$$

Here the property of independence is employed so that the joint probability is the multiplication of the two marginal probabilities. Again, this recovers the form of  $p_0(\mu)$  in equation (14). Thus the discrete reaction index distribution also remains the same.

As a sub-conclusion, one can see the reaction probability density undergoing within the two Gillespie Monte Carlo methods remains the same, which means these two methods are equivalent.

### 3 Simulation Result

To test Gillespie Monte Carlo algorithm, a simulation for the dynamic system shown in figure 1 has been done and the results are displayed in figure 5 and 6.

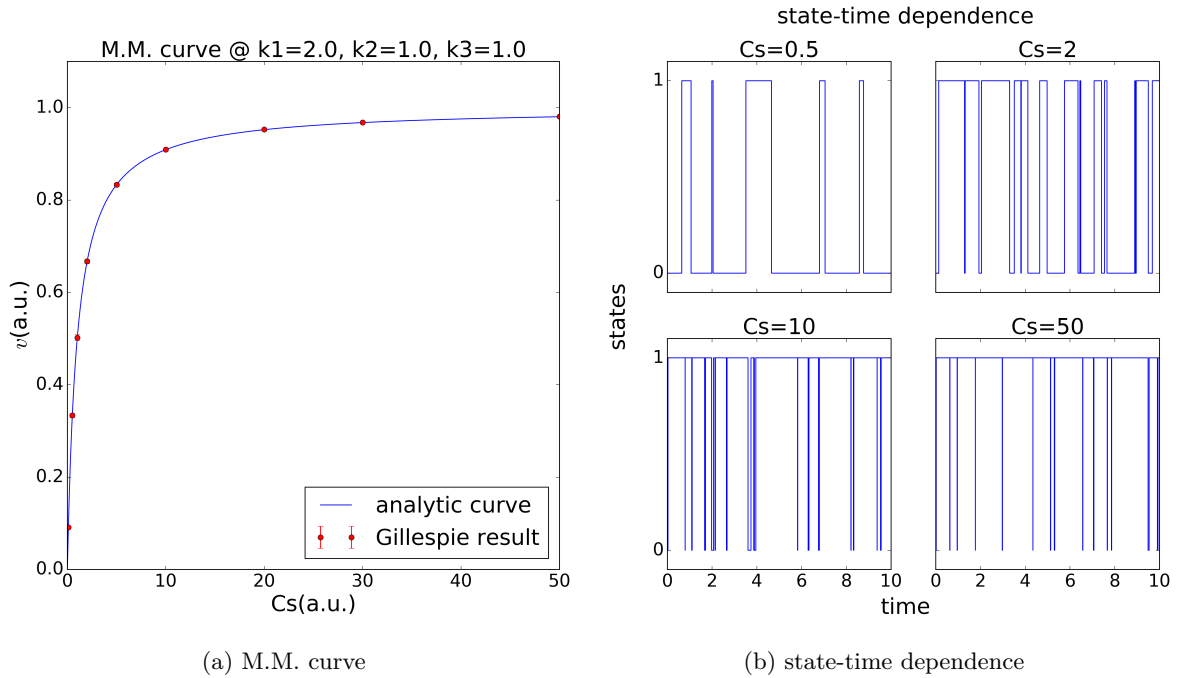


Figure 5: simulation result of direct Gillespie Monte Carlo method

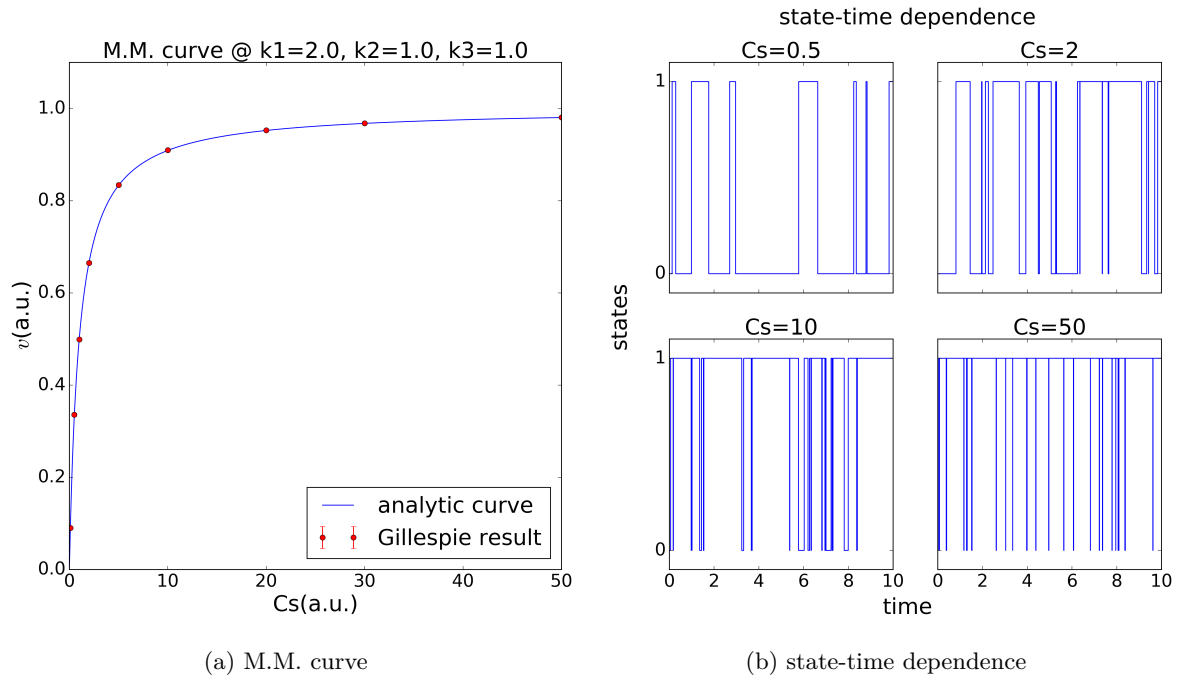


Figure 6: simulation result of "first-reaction" Gillespie Monte Carlo method

10,000 Monte Carlo steps have been run for each substrate concentration point. The errorbars are obtained by computing the standard deviation of each point. As can be seen, both of the two methods function well and a 10,000-step simulation is enough for the system shown in figure 1 due to the very small deviation. The state-time dependence is plotted out as well for four different substrate concentrations. Clearly, as the  $C_s$  increases, the average life-span time of state 1 increases, which is consistent with mechanism described by the chemical equation (1).

In conclusion, the equivalence between the direct and "first-reaction" Gillespie methods is shown by both analytic deduction and simulation results. In addition, the simulation results of reaction rate are very consistent with Michaelis Menten equation. For the simple system shown in figure 1, 10,000 Monte Carlo steps can already give fairly small deviation, while for a more complex system, more might be required to lower the deviation. Because it is easy to program and the error is easy to be controled, Gillespie Monte Carlo algorithm is potentially a powerful method for studying various biochemical systems.

## References

- [1] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.
- [2] Daniel T Gillespie et al. Exact stochastic simulation of coupled chemical reactions. *J. phys. Chem*, 81(25):2340–2361, 1977.
- [3] Terrell L Hill. *Free energy transduction and biochemical cycle kinetics*. Courier Corporation, 2004.
- [4] Leonor Michaelis and Maud L Menten. Die kinetik der invertinwirkung. *Biochem. z*, 49(333-369):352, 1913.